AI and Ethics
https://doi.org/10.1007/s43681-021-00110-y

ORIGINAL RESEARCH

Disruption, technology and the question of (artificial) identity

Dina Babushkina[1] · **Athanasios Votsis**[2]

Received: 11 August 2021 / Accepted: 4 October 2021

Abstract
The current state of human–machine interaction has set forth a process of hybridization of human identity. Technology—and most notably AI—is used as an effective cognitive extender, which enables the extension of human personhood to include artificial elements, leading to the emergence of artificial identity. Discussing—and accommodating—anthropomorphization in human–machine interaction should no longer be the primary focus. Rather, the scope and quality of frameworks in which the hybridization of human identity occurs and evolves has significant ethical implications that pose very pragmatic challenges to users, the industry, and regulators. This paper puts forth a few main principles upon which such a discussion should evolve. We illustrate why disruptiveness can easily turn into human harm when the frameworks facilitating it overlook the human vulnerabilities that arise from hybrid identity, notably the asymmetric and asynchronous relationship between the human and artificial counterparts. Finally, we claim that these new types of vulnerabilities, to which a person is exposed due to the intimate degree of pairing with technology, justifies introducing and protecting artificial identity as well.

Keywords Anthropomorphism · Artificial identity · Cognitive extenders · Disruptive technology · Ethics of technology · Smart devices

1 Tech, disruption, and the human condition

The domain of human–technology interaction is a complex one and can be approached from multiple angles, each with its own questions and priorities. Our starting point is the question about the human condition and the need to understand, first, how the involvement with technology affects personhood and, second, to what extent these effects influence the technology itself.[1] Within this broader question, we are interested in the question about identity and how the concept of identity transforms due to certain forms of involvement with technology, leading to a strong degree of cognitive pairing. Such pairing promises to be at its strongest in the case of AI-based technologies, due to their natural propensity to emulate, supplement (and often substitute) human cognitive skills. We wish to draw attention to a few distinct conditions that make modern technology disruptive. The goal, however, is not to conceptualize disruptiveness as such. We will not develop a generic account, looking for the abstract principles that are applicable to any technology; similarly, we will not develop an inventory of potential radical changes that technology may produce in social, political, or economic domains.[2] The general goal is, rather, to identify the disruptive effects technology may have on personhood: What critical effects can it have on personality? How do these, in turn, affect our conception of agency and identity?

✉ Dina Babushkina
d.babushkina@utwente.nl

Athanasios Votsis
a.votsis@utwente.nl

[1] Section of Philosophy, Faculty of Behavioral, Management and Social Sciences, University of Twente, Enschede, The Netherlands

[2] Section of Governance and Technology for Sustainability, Faculty of Behavioral, Management and Social Sciences, University of Twente, Enschede, The Netherlands

[1] This is a unique contribution of this paper to the discussion about disruptive technology, which mostly focuses on its socio-economic effects (see e.g. [10, 18, 26, 32, 49]). A line of research that is closer but not directly relevant to this paper is conducted by Nickel [38], who, following Baker [4], understands disruptiveness as undermining established moral norms and creating moral uncertainty.

[2] A systematic work in this direction is currently being undertaken by the research programme Ethics of Socially Disruptive Technologies realized by a consortium of Dutch universities. For a comprehensive overview of the potential disruptions that technology may cause in society, see [7] and [30].

Published online: 25 October 2021

Some clarification about our understanding of disruptiveness is needed, however. In this paper, we will work under the following assumption.

> For some phenomenon A to be considered disruptive for another phenomenon B, A's effects on B, direct or indirect, must be such that B is not able to continue to exist and to transition smoothly to another state, given the rules governing B's nature.

On one hand, this means that when B is experiencing disruptive effects by A, B is either forced to undergo a rapid radical change (i.e., incompatible with a smooth transition) from one state to another, or faces a threat to its very existence. There are cases, however, where the disruption does not occur because of the technology itself. Here, A does not present an immediate threat to the integrity or existence of B, but, on the contrary, A promises to facilitate B's smooth transition from one state to another by creating favorable conditions for the development of some of B's predispositions. In such cases, the framework[3] within which this transition occurs, is lagging so much behind, by constraining and/or distorting the transition, that it is jeopardizing the conditions which guarantee that B's change will be uninterrupted and harmonious, therefore, effectively rendering the transition disruptive. This is important: when having the goal to make technology safe, equal attention must be paid to this second factor to identify and mitigate risks that are associated with outdated frameworks. This paper aims to point to an area in human–technology interaction (HTI) where the disruptive effects of technology are conditioned by the lagging framework, instead of the technology itself.

To sum up, the intention here is to identify risks that the use of modern technology creates for personhood and the concrete vulnerabilities of the agency in the light of the use of technology. In realizing this intention, we propose to shift the attention from the problem of anthropomorphization of technology to the problem of hybridization of agency (human-technology hybrids) and explore the disruptive potential of technology for the cognitive practices of a human. At the same time, we want to understand how this process affects artificial agency: how does pairing with a human affect artificial agency? We believe that in this domain we can find tools for a conceptual change, i.e., ways to rethink not only human agency but to conceptualize artificial personhood as well. This will allow us to answer an ethical concern about minimizing the risks involved and lower the vulnerabilities of the users.

2 Shifting focus away from the anthropomorphization problem

Anthropomorphization has become somewhat of a commonplace discussion in the fields of human–computer interaction (HCI) and human–robot interaction (HRI), resulting in rich literature on the topic (see e.g., [1, 17, 19, 39, 60]). And it is not surprising that those who are worried about the effects of technology on the human condition tend to look for the disruptive effects of modern technology in the phenomenon of anthropomorphization. According to our definition,

> A person may be said to anthropomorphize an object iff (a) he/she is attributing to it typically human (or, broadly, animistic) features (including conscious and unconscious beliefs, as well as emotional reactions) and (b) guides his/her actions towards it by the belief that it has such features (including treating it as being animate and expect it to respond to such treatment in the manner appropriate for a human).

In other words, there are two elements to it: a cognitive (a belief that an object/artifact has certain human or animal-like abilities) and a behavioral one (acting upon this belief). The narrative about the dangers of this sort of stance in HTI goes something like this. Modern technology, such as personal computers, social robotics, and what is commonly called "smart" devices,[4] has reached an incredible level of sophistication. Their ability to interact with their environment and human agents in a manner that gives an impression of spontaneity (in the manner that is not determined by a human while the interaction is happening), alongside their ability to adapt to the user by collecting and analyzing complex data about him/her on the fly, justifies the unique status of their artificial agency. This further allows them to imitate and even substitute what has always been considered solely human functions. Of course, this is taken advantage of in many social spheres where various types of devices are used to substitute or supplement persons in providing various social services to humans (e.g., care, medical help, or therapy). But the substitution does not end there, as AI technology is being integrated into relationships between individuals, such as friendship, love, or care. AI-based systems—either embodied or virtual ones—are expected to perform the same functions as human partners in such

[3] By 'framework', we understand broadly all the relevant systems within which the interaction between technology and humans happens, such as the conceptual apparatus, scientific paradigm, legal system, governance, social organization, moral norms and ethical theories, as well as the norms governing tech-related professions.

[4] We use "smart" devices in a broad sense, referring to technology designed to engage and interact with the user, and utilize obtained data to adapt to the user.

relationships. This includes, among other things, providing certain—for the lack of a better word—services to the user, of the kind that we are morally justified to expect human agents to provide us, in virtue of them standing in a certain type of relationship with us. Showing affection, comforting, listening to a friend who experiences distress, or even such a simple thing as asking how your spouse's day has been.[5] As a result, researchers raise a worry that the very nature of the interaction between a human and technology (including, the design of a device/program, its marketing, the narratives about devices, and our practices of interacting with them) invites, encourages, and reinforces anthropomorphization of our devices. The disruptive potential of this phenomenon for the human condition is multidimensional, but, perhaps the most fundamental risks are the degradation of relationships between persons, frustration of rationally justified expectations and subsequent psychological distress, and the objectivization of persons.

First, treating artificial agents as identical to human partners in interpersonal relationships threatens to downgrade our understanding of the very nature of interpersonal relationships. The asymmetry between human and artificial partners (e.g. a human lover and a robotic lover) is apparent in the inability of artificial agents to reciprocate (e.g. to feel affection towards you or be concerned with your well-being, see [56] and [2]) and, as a result, their inability to meet the expectations of certain types of reactions, behavior and the ways they treat us when such expectations are justified and sometimes required in virtue of the type of relationships that bond the two partners. Or, rather, they would be justified and sometimes required if both partners were persons. Given the conceptual asymmetry between a human and an AI-equipped robotic partner, there is a need for adjustments that would provide a better fit for the expectations. Anthropomorphism predisposes for the following route: change the concept corresponding to a certain type of interpersonal relationship (such as love) in such a way that it would also include artificial partners and accommodate their limited capacities to reciprocate (see e.g., Levy in [58] and [34]). For instance, Turkle [56] and Richardson [43] warn that this will lead to the degradation of relationships between human partners or even make them unattainable to those who require reciprocity, because technology would make them redundant. The inclusion of artificial agents into the realm of partners in interpersonal relationships will lead to the loss of certain elements of the meaning of such relationships, loss of depth as well as atrophy of practices and attitudes that we consider to be valuable and, perhaps, even essential to our well-being. In short, the risk is that to be able to have more human interaction with artifacts, we will end up having more superficial relationships with humans.[6]

Second, on the personal level, this in turn has a potential to cause psychological distress. The human agent might have an acute need in reciprocity which will remain frustrated (see e.g., Evans; Boden in [58]). This may be worsened by a certain cognitive dissonance, i.e., distress from experiencing contradiction between beliefs and values, when taking part in actions (or standing in a relationship) that contradict at least one of them. The risk of finding oneself in the state of cognitive dissonance comes from the contradiction between the agent's own belief about certain aspects of interpersonal relationships (such as reciprocity) being valuable and the implicit devaluation of this aspect in interactions with the artificial agent. This may in the end lead to something that a Hegelian would call "alienation" from oneself, i.e., losing connection with oneself and one's psychological needs.

Third, on a more abstract level, anthropomorphization of artificial agents leads to objectivation of persons, i.e., to the reduction of the persons to object, both in the ways we think of and treat other people. The worrisome part of this process is the loss of respect to personhood and devaluation of its essential characteristics. Richardson [43] explores this risk in application to sex robots and the ways in which they may lead to the objectification of women. Babushkina [3] argues that engineers have a moral imperative to integrate respect to personhood into social robotics.

We do agree that these are serious issues that need attention, but we also think there is a need to look beyond anthropomorphization, which is just one part of the story. It highlights certain problems but does not include into its scope others (cf. e.g., [16, 48]). Our worry is that anthropomorphization is becoming somewhat exaggerated in HRI and it takes attention away from other processes and newly emerging phenomena that are at least equally important and harbor problems that need attention.

What is important to note is that anthropomorphization is a much broader phenomenon documented across human cultures and domains that humanities have long been dealing with [6, 24, 25]. Transferring human features to technology is just one instance of treating inanimate things as if they were alive. Merely referring to the tendency of anthropomorphization does not capture the nature of the interaction

[5] A broad range of such uses is discussed in [58].

[6] Nyholm and Frank in [15] look into the conditions under which a robot could possibly love a human who sees her/himself as being in love with the robot. Their conclusion is that it is only possible if robots achieve an extremely high degree of sophistication, e.g. be capable of certain commitments and value-beliefs. The article is also interesting, because it discusses what types of commitments a love relationship between persons presupposes, and to what extent these should play a role in deciding whether a robot can be seen as exhibiting love. Danaher [14] looks into another type of interpersonal relationship, friendship, and argues that robots can fulfill the virtue-ideal of friendship and that they can perform certain friendship rolls, and even enhance friendship between humans.

with technology as a unique entity, and the problem itself does not grasp what could be considered a uniquely technological impact. The existing toolkit of humanities should be sufficient to address the issue and provide sound arguments against anthropomorphization in the HRI/HCI domain.[7] Indeed, as a matter of fact, this is what we see in HCI and HRI literature[8] when it comes down to the most essential discussion, that is the normative implications of the human tendency to anthropomorphize technology. In HRI and HCI, it is often taken as a fact that anthropomorphization has normative implications which have to be taken into account during the design of robotic and AI systems. It is not uncommon to see this sort of implicit reasoning: since people tend to attribute human features to their devices, get attached to them as they would to humans, and do treat them as they were in some sense like humans, therefore the design process should meet this demand and we should construct devices with more human-like identity. Of course, often the motivation behind this is making devices easily acceptable by potential users, predisposing them to interact with the devices on a broader and deeper scale (what is commonly referred as "trust" in the device). Classic psychological theories, on the other hand, would see in anthropomorphization a degree of psychological immaturity (as we can see on the example of animism, i.e. the tendency to see natural phenomena as possessing such properties as will), or perhaps even disorder (e.g. delusions), and philosophers would categorize this as a case of a mistaken belief.[9] Therefore, if anthropomorphization itself is a problem, then the way to tackle the problem is to correct the belief about what artificial agents are and what the limit of their capacities is. Just in the same way as you would address—if there is such a need—one's belief that a stone statue of a god may be angry with someone, or that an amulet may protect from harm. They are just not those kinds of entities: a stone is an inanimate object whatever shape it receives and it cannot feel or express any psychological states; the amulet is an inanimate object as well and, therefore, does not have the capacity to act or produce any motion on its own.[10] Changing the user's beliefs about the nature of artificial agents, and other non-agential devices they are interacting with, can only be achieved by adjusting our concepts, refusing fantastical and metaphoric narratives about technology and switching to factually informed narratives to encourage users to form more realistic expectations about their devices. In other words, the imagery of strong artificial intelligence, as well as Turing's general learning agent, keeps creeping into the current state of smart technologies—which are certainly still in the domain of weak artificial intelligence—and malinform the design process. Thus, a necessary element here is informing design and marketing in such a way that does not feed illusions of reciprocity, instead of trying to accommodate those unrealistic—and rather disruptive—notions into the very design of the product.[11]

We will not, however, go into more detail about this. We only want to attract attention to the fact that the disruptive effects that technology has on personhood and interpersonal relationships due to the user's tendency to treat artificial agents as if they were human are not necessitated by the nature of HTI that we observe. They are just yet another expression of animism, i.e., a result of a well-documented psychological tendency of people. As a result, despite being provoked and encouraged by technology, these dramatic effects are not caused by or are unique to it.

3 Hybridization and cognitive pairing

We have argued that overconcentration on anthropomorphization drives attention away from some of the emerging HTI processes that raise ethical concerns. One of these processes is the hybridization[12] of personhood.[13] By this, we understand

> The situation when the capacities and properties that are commonly thought as determinants of personhood become merged with technology, in one way or another.

This problem, we believe, reflects more accurately the changes in personhood that happen due to the interaction with technology (as a unique entity, distinct from other objects and natural phenomena). This does not happen, because the user transfers a certain pattern of perception or belief on a device. For a strong dependency to occur, you do

[7] Given that anthropomorphism is not the main topic of this paper, we are unable to go in detail into the discussion of history of and main theories about anthropomorphism. This discussion is a matter of a separate research paper.

[8] For a literature review on anthropomorphism in AI-enabled technology, see [35].

[9] See e.g. [13] on anthropomorphism as a cognitive bias.

[10] This can be seen as a case of *a category mistake* when a representative of one category of entities is treated *as if* it was a representative of another.

[11] Leong and Selinger's [33] taxonomy of dishonest anthropomorphism might be interesting to the reader in this respect.

[12] We understand hybridisation in broad terms here, as a phenomenon where human and artificial are in some way acting as a whole.

[13] One topic discussed in research literature, which is broadly connected to hybrid personhood, is that of hybrid agency. However, this topic focuses on a different host of problems. It primarily deals with the question of action and decision making. For that reason and given the goals and scope of the paper, we will leave the discussion about hybrid agency out.

not need to anthropomorphize your tablet and your laptop. You may develop a strong bond to your connected devices, or even pair with them, without any shadow of belief that they are more than what they are: inanimate artifacts. The problem we want to discuss emerges as a natural consequence of the interaction with the device: we are faced with such intimate pairing between a human agent and technology in the process of solving cognitive tasks, that changes in the technological component of this pairing have a disruptive effect on the agency of the human.

Thus, our point is that it is not so much the case of misattribution of human features to technology that is most informative for the understanding of the human condition in the context of HTI, but the co-dependency of human and artificial agents for the formation of their identity. When we shift the attention to the hybridization of personhood and acknowledge the degree of the investment of the user's personality into a device, we will be able to better appreciate the uniqueness of modern technology (especially the various "smart" devices, see e.g., [5]) and its true effect on personhood. Let us take an example of conversational agents, i.e., artificial agents that—due to natural language processing (an AI technique)—allows direct verbal communication with the user (e.g., chatbots or virtual agents). Due to the desire to increase their usability, much of the discussion concerning conversational agents evolves around their human likeness (e.g., as avatars, digital twins, robots), with the main philosophical intrigue being the acceptability of encouraging the tendency of users to treat various conversational agents (functioning as personal assistants, health chat bots, digital companions, therapists, etc.) as is if they were humans. Be as important as it may, the focus on the human-like appearance and human-like behavior of the artificial agent is missing an entire host of issues that stem from the fact that human agents do not just interact with the artificial conversational partners, but merge with them, whether they appear human-like to them or not. This merging or pairing can take different forms. Through the language-processing algorithm, the user's data decides what the artificial conversational partner is, while the algorithmic determinants set constraints on the user's cognitive processes and experiences which constitute a conversation. This affects the functionality and utility of conversations, especially when these happen between human agents, mediated by an algorithm. The extent to which the nature of the conversation changes and the implications of the human–computer pairing for conversation as a cognitive practice as well as for cognitive agency, are important questions with high ethical relevance and they are independent of the appearance of the technology for the user.

In a nutshell, we need to take a closer look to the fact that modern computer technologies, especially based on AI, function as *cognitive extenders*. Cognitive extenders fall under the category of non-autonomous systems, which means that they do not perform tasks on their own, but function as aids for the completion of various tasks by humans. A cognitive extender is

> "[A]n external physical or virtual element that is coupled with the human to enable, aid, enhance, or improve cognition, such that all—or more than—its positive effect is lost when the element is not present" [29].

Several things need to be noted about this definition. First, by locating extenders outside the physical brain of the user, this definition sets extenders apart from cognitive enhancers, such as nootropic drugs. However, it does not draw a sufficiently precise line between extenders and other types of enhancers, which do need to be physically incorporated in the brain. Any extender can become an enhancer if it enables its user to perform a certain cognitive task on a level beyond what is normally possible for the human brain. In other words, if it is the case that when an extender is removed, all that a person has lost is a competitive advantage, then it was used for enhancement. But what is more interesting, is the distinction between extenders on the one side and tools offering "cognition as service", i.e., a range of tools that "augment and scale human expertise", increasing "productivity and creativity… with the help of cognitive assistants" ([53], see more on cognitive assistants [37]). The difference here lies not so much in the design of a specific device or its functionality or purpose, but in the degree of pairing between it and the human agent when it comes to performance of a certain cognitive task.

It is tempting to reserve the term cognitive extender for cases where a device helps to restore or substitute an impaired cognitive function in its user (assistive technology).[14] Some of Vold and Hernández-Orallo's [57] examples of AI extenders fits this narrow definition. One example is that of Helen, an elderly lady with Alzheimer, who relies on augmented reality glasses that help her to orient herself in her environment, performing some of the cognitive functions for her: they help identify objects and persons, classify situations she finds herself into (for example as potentially dangerous), and plan her day. Another example is that of Lewis with ADHD (attention deficit hyperactivity disorder).[15] He uses a special AI device that keeps track of his activities and brain functions, produces recommendations and trains certain skills in a game manner. But even if such an AI device would have been unavailable to Lewis, he could have equally aided his cognitive tasks (such as attention, focus, planning)

[14] More on the use of intelligent assistive technology for dementia see [31].

[15] Other fields of application of AI extenders in the field of mental health are: addiction, borderline personality disorder, and autistic disorders (see 57).

with different technologies: e.g., with a combination of various applications on his laptop, tablet, and/or mobile devices. Calendar programs, project planners, reminders—when these are carefully tuned to Lewis' needs—will help him to function normally when it comes to structuring his day, staying on track with studies and planning other activities. There are other applications that can help stay focused on a task. Sometimes, it can be as simple as arranging a certain type of background music/noise. For those who, like Helen, are suffering from memory problems, a solution may lie in combination of memo application, notebooks, photo programs, carefully selected and personalized for easy and swift use throughout the day. Even a paper notebook can play the role of a cognitive extender if it functions as an analogy to biological memory[16] (cf. [11]).

However, it is important to note that Vold and Hernández-Orallo's definition of cognitive extenders is designed to accommodate a much broader spectrum of cases, that is when essentially no impairment of cognitive capacities takes place. What is crucial is whether there is a sufficient degree of pairing between it and the human mind, that is, when the role of an external device for the performance of a certain cognitive task is too great and viewing it as an optional tool no longer makes sense. From this standpoint, my laptop (tablet or a mobile device) functions as my cognitive extender, because there is no longer an easy way for me to draw the line where my own cognitive contribution stops and the application on my device starts when it comes to such activities as recalling things, planning my day (both work and hobbies), spelling and word processing, creating narratives, navigation, and even selecting where to direct my attention, and what information is relevant. And all this with no AI involved, only via the interaction with an ordinary computer. Some 10 years ago, perhaps, it would have been easier to draw this line, and for some people it probably is. However, nowadays, the degree of reliance on computer technology for the whole range of cognitive tasks is so great, that separating them is no longer possible without a substantial loss in productivity and comfort of work and life. One cannot help but agree with Hernández-Orallo and Vold saying that "[o]ne very interesting feature of these interactions is the way the user changes their reasoning processes: it is not that part of the process has been replaced; rather the whole task has been redesigned, and the skills of the human user often coevolve with the technology" [29], p. 507).

This broader definition of cognitive extenders has great potential for inquiring into the disruptive potential of technology for human personhood, and at the same time, holds the key to the solution of how to mitigate the devastating effects. First, it allows us to stop seeing the process of HTI as involving two independent entities: a human agent and an artificial agent. We move towards perceiving them as hybrid agencies, i.e., as unique unities of distinct elements, human and artificial, which are interdependent. This allows to re-conceptualize the HT pairing in such a way that is inclusive of various degrees of integration and co-dependency between these two participants. This in turn, allows us to recognize new types of vulnerabilities that human persons obtain as a result of entering the dependent relationship with an artificial agent, and to validate the negative experiences that the users suffer or may suffer as a result of their vulnerabilities (mitigation of disruptive effects associated with this type of technology). A major source of vulnerabilities for personhood lies in the underresearched fact that in the process of pairing with the device, the user invests her/his personality (and often life) in the piece of technology. Because of this, certain types of alternations of such a piece of technology have a disruptive effect on the user. This is not a minor thing: bringing forward the discussion about the dependency of human personhood on technology, prepares the ground for ethical reasoning justifying certain changes and policies. An example that illustrates this is as follows. Alan, a while ago, started using a certain combination of hardware and software, say a laptop of a certain architecture that is paired with an appropriate operating system. He has developed an ecosystem of productivity software that is working quite well, say an ecosystem of calendars, note-taking, and task reminder software. It has been working well, and Alan has integrated this in his workflows, while investing a significant volume of his data and information in his devices. However, the rolling update framework and periodic major release framework of many elements of his ecosystem one day cause his information and ecosystem structure to be lost. Some elements do not work any longer and, although some data are recoverable, the previously established pairing is lost. This has significant ramifications for Alan's productivity, cognitive performance, and well-being that cannot be assumed away by the usual thesis that all information is recoverable and all hardware is replaceable. Paraphrasing Simon [51], the whole is more than the sum of its (replaceable) parts, and serious problems arise when one is faced with disruption of that whole (the hybrid identity) due to the framework governing the updating of its parts (currently delegated to the industry). We elaborate on this next.

4 Extension of personhood and artificial identity

By becoming the locus of certain elements of human cognition, devices effectively extend human personhood. The relationship between the device and the human mind is no longer the one of instrumental type. The user relies on the functions of the

[16] On the use of digital systems and services as memory extensions see, e.g., [12] and [52].

extender the same way as she/he would rely on her own cognitive skills. The significance of the artificial agent for the human cannot be reduced to that of a mere tool for the solution of a certain goal at hand, at least not without an interruption and change in the continuum of the user's cognitive practices, that is not without a damage to the identity of the user. As a result, if the separation from such a device or its alteration occurs, it is not surprising that the person may no longer feel the same or capable of dealing with certain life and work situations in the same productive manner, experience loss of certain abilities, feel a vacuum and face the need to re-learn how to solve familiar problems, or even develop a new set of skills. And again, this is true of even the very mundane technology, such as our laptops and phones. Just imagine losing your laptop, and the backup hard drives; or your cloud storage being compromised.

Now, how can this be explained? Why do cases like unsolicited external tampering with the artificial component of the human–technology hybrid agent have such a disruptive potential for human personhood? And why, again, it is not enough just to substitute the lost device with a similar one? What significance does our devices' *staying the same* have for our personhood?

These questions point to another side of the hybridization phenomenon—the emergence of artificial identity, necessitated and warranted by the propensity of the human person to stay the same, i.e., continue to retain identity within and despite the merging with an external device. Through their interaction, the artificial and human agents develop a unique psychosynthesis and identity co-dependency: to stay the same each component relies on the other to remain the same. This leads to a vulnerability of human personhood to changes in the artificial component. And this, in turn, ethically justifies protecting the artificial component from such changes, which, in fact, amounts to a requirement to preserve its sameness. To do that, we have to conceptualize artificial identity.

How are we to approach artificial agency in the given context? Arguably, any piece of technology is (a) reproducible, i.e., exists (or can exist) as more than one token, and (b) replaceable, i.e., any individual token can be substituted by another or similar copy. This, by definition, makes them not unique. The situation is getting even more difficult for "smart" devices. They are designed in such a way, that (a) allows a greater interchangeability of hardware (i.e., not only by same or similar copy, but by a completely different physical device), and (b) allows the same instance of software (preserving all the user settings and data) to exist on multiple hardware on the same time. Therefore, if the question of identity is one about the conditions for staying the same in one's unique characteristics, then where does the unique identity of each token start?

The problem is that not all tokens will have an identity. Many will remain nothing more but replicas. Therefore, what decides whether a given token (say, my tablet vs. yours) will come to have an identity? We suggest that it makes sense to talk of artificial indefinity when the actions that alter or destroy the device, count—from the perspective of the user—as a non-trivial change or a disruption in his/her life. Throughout the history of her/his interaction with such technology as a laptop or a robot, the user has invested a part of her/his own personality into it, by fine tuning the settings, training algorithms with her/his own data, interconnecting various applications and servers due to her/his specific needs and preferences, feeding information about events in her/his life to various systems, and making all that influence, structure, assist, and often regulate various aspects of her/his life, work, and interactions with others.

Of course, in principle, this is true of any piece of technology. Even an artist's chisel may come to possess identity if it becomes her/his favorite tool, extension of his/her hand and mind. It is unique, because it cannot be easily substituted with a new one. Even repairing the instrument may destroy its unique properties that the artist relies upon to deliver the results that she/he expects. The instrument is unique, because it is finetuned to the artist and her/his unique needs and abilities. Without it, the objects created would not have had the same distinct features, would not bear the mark of the craftsman, and, which is equally important, the artist would not have had the same quality of the creative experience. But computer technology, and even more so AI and "smart devices", not only have a stronger tendency for that, but also altering them tends to be more disruptive for the user. This is due to the fact that these, more often than other types of technology, become the extenders of the user's personhood. However, the locus of artificial identity remains, as Hegel would put it, "for-us". This does not mean that it exists only in the imagination of a human. Artificial identity is constituted by tangible elements, such as specific settings, modifications of the system, memory, etc.[17] But what we mean by "for-us" is that without the user, whose cognitive capacities it extends, the artificial identity would be meaningless. Staying the same

[17] An interesting question has been raised by a reviewer: since one of the general goals of artificial intelligence as a field of computer science is to build an artificial identity that combines attributes of multiple humans in one super-agent that will outperform humans in various tasks, where does this leave us with respect to the problem of biases in machine learning? Within the context of this paper, the phenomenon of biases in machine learning can be seen as the direct consequence of the hybridization of cognitive processes. It signals the fact that this pairing may (and, more often than we hope, will) happen in undesirable ways. As a result, we have to discuss in more detail the desirability and acceptability of certain types of human-AI pairings as well as the possible limitations on the hybridization in order to prevent tuning artificial identity in a harmful way. Unfortunately, we cannot explore this topic in any length in this paper. The question "What the (dependent) artificial identity should be?" certainly deserves separate research.

in its unique characteristics is conditional upon the imperative to prevent disruptions in the cognitive life of a person. "For-us" also means that devices themselves are neither aware of their identity, nor have a need to exist in one form or another, or to protect a certain aspect of themselves from change or destruction.

What we find interesting about technology is that it is often misleading to talk about it in terms of general principles or a priori qualities. The true intrigue is in conditionals, in concrete and tangible details of user experience. It is the unique conditions of the specific human–technology interaction environment that creates and shapes the identity of a device. These specific conditions determine whether the device remains an interchangeable copy or becomes a non-disposable, non-interchangeable, unique entity. This brings us to our main claim: artificial identity emerges in the interaction of the device with the user, and it is through this interaction that the device obtains its distinct features and the locus of its identity. Artificial identity, according to this view, consists of the extension of its user's personality, as a synthesis with the user's phycological states. It is in this psychosynthesis that a device develops its unique characteristics.

5 Artificial identity: why not just another person?

We have proposed, essentially, a dependent concept of artificial identity, i.e., a view according to which the identity of an artificial agent exists for a human user and is justified by the fact of the intimate integration of technology with human personhood. There are multiple angles from which one could address the question about artificial identity. One—and perhaps the most expected in the philosophical context—way to approach this issue would be to consider whether to ask whether an artificial agent can in principle be considered a person[18] (on different aspects of this question see e.g., [8, 20, 22, 27, 42, 47]). The reader should not, however, expect from us a general philosophical discussion about the nature of personal identity—it will not be very useful considering the paper's goals. Formulated this way, the question about artificial identity takes us in the hypothetical realm, where we will need to investigate the assumptions about future technology, i.e., a type of AI (strong AI) and properties which it may or may not come to possess. We do not deny that such discussion is scientifically interesting and important, but it simply falls outside our scope since we want to focus on existing technology and the effects it already has on human personhood.

We do, however, want to give our reason why we believe that artificial identity, as described in this context, should not be conceptualized analogously with personal identity or, to that extent, to the identity of animals (on the latter see, e.g., [21]). We give two reasons for this: the asymmetry and asynchronicity between personal and artificial agency compared to what can be called their "life cycles", and the fact that artificial identity presupposes a non-derivative right to persist.

The first reason is that constructing the notion of artificial identity on the basis of comparison with personal identity is bound to be asynchronous and asymmetric. Generally, asynchronicity refers to non-correspondence between the temporal locus of human and artificial identity, while asymmetry refers to non-correspondence between their spatial loci. Taking asynchronicity first, if we were to consider the "life cycle of an identity", i.e., the period from the time when the entity can be said to start being a unique individual to the point when it is no more, we are bound not to find the correct correspondence between humans and artificial agents. One reason for that is that there is generally no agreement about the beginning and the end of personal identity, while artificial agents (hardware or software) are much easier to tokenize or segment and define temporally. Turning to asymmetry now, the fact is that artificial agents have unique properties which cannot be unproblematically transferred to human agents, and vice versa. One of these unique properties is the high degree of independence from embodiment: the software, containing all essential unique features of an artificial agent, can easily be detached from the hardware containing it, transferred to another hardware or exist on multiple devices simultaneously. Consequently, the elements of artificial identity are of a different nature than those of human identity. Hybridization of identity challenges not only the expectations of a synchronous and symmetric relationship between human and artificial identity, but it also challenges the inherent such properties in both identities alone.

DiGiovanna [21] argues that there is a need to take into account "the increased malleability of personal identity that this technology affords: an artificial being can instantly alter its memory, preferences, and moral character". This ability widens the divide between artificial and personal identities even further. The consensus among philosophers is that personal identity is a matter of some sort of continuity, be it psychological (see e.g., [36, 41, 50]),

[18] Please note that this question, even though related to is nonetheless different from such questions, as (1) Whether artificial agency is similar to human agency; and (2) Whether artificial intelligence can be considered analogous to the human brain.

physical [40, 55, 59] or narrative (e.g., [44–46, 54]).[19] The capacity to overwrite oneself is, thus essentially a "personhood-defeating capacity". Compared to personal identity, artificial identity is much more flexible and capable of instant change, i.e., "reworkable". This unique ability of the artificial agents or, as DiGiovanna [21] calls them, "para-persons", calls for a change in our concept of moral agency (as a consequence of being a person) and for the readjustment of our moral judgements so that they can also fit the unique capacities of para-persons. Furthermore, across a wide range of capacities and characteristics, artificial agents are able to achieve perfection on the scale and speed which is not available to humans, and thus could develop into supra-persons [23].

The second reason for which we have to refuse constructing artificial identity analogously with personal identity is the following. The ascription of identity to a human or animal entails something that we will call here the non-derivative right to persist, i.e., the right to stay the same (continue to exist as the same entity) and not to be subjected to changes significant enough for them to stop being the same. In other words, animals and humans enjoy the right not to be mutilated physically and altered psychologically. And if that happens, these actions should be condemned as torture. Humans and animals have such a right simply in virtue of having their unique identities (not in a legal, but rather in a moral sense). Objects, on the other hand, only have the right not to be significantly altered iff they are property (but then again, I am free to do anything I wish to my property, but not someone else's) or of recognized value (such as art objects). Thus, the ascription of identity to objects is crucially different. Now, where does artificial identity belong? Devices do not have the non-derivative right to persist like animals and humans do. This is clear from the ease with which devices are modified, updated, or exchanged for newer models. By default, the software in our computers, tablets and phones is regularly updated in the manner that significantly alters interface, functionality, personal settings, and as a result overall performance and workflow.

Why, then, not to give up the talk about artificial identity altogether? The reason is that artificial agents, while intimately paired with human agents, are physically embodied independently from the user's brain or body. They are, in one way or another, instantiated in the world, as distinct and destructible tokens. The set of vulnerabilities that human personhood is open to when conditioned only by its body and mind is different from the vulnerabilities that personhood has when conditioned by a combination of body, mind, hardware and software. One could now hurt one's person without doing anything directly to the person's mind or body, but by just attaching the hardware or software of the extender. The human person becomes also vulnerable to types of harm that had not been applicable to him/her before such pairing, i.e., to threats of altering or terminating functions of the device they are paired with. In a nutshell, certain aspects of personal identity become dependent on the sameness of the unique characteristics of devices, with which they are sufficiently paired.

This makes it clear why there is a need to conceptualize artificial identity, even despite the fact that the concept of personal identity appears to be a poor analogue. The approach we have outlined has a good potential for creating such an alternative account, which will help to (a) identify the right conditions under which identity can be ascribed to a device/piece of software and (b) conceptualize artificial identity uniquely as an identity of a device (that is of a piece of technology as opposed to a human or an animal), respecting its ontological status and epistemological role.

6 Ethical implications

The account of artificial identity outlined in this paper poses a serious challenge for developers and existing practices of software lifecycle such as update schedules. Such practices will have to be modified so that they avoid unsolicited changes in the devices which entail disruptions in the functioning of the user.

This is why. Even though, artificial identity is to be found in the unity of the artificial and human agents, this unity is such that (a) only the human is interested in persisting in the future as well as in not being harmed, and (b) there is a marked disproportionality of the harmful consequences in case the integrity of the hybrid agency is threatened. Because the user has invested her personality into the device, what threatens artificial identity exposes the vulnerability of the user. By pairing with the device, investing her personality in tuning the device, she renders herself vulnerable to the risks of harm in case the device stops being the same. This, as we discussed above, makes artificial identity, in effect, an extension of human identity, and provides it with a moral status (cf. [9, 28, 29]). While the mutual vulnerability grants the identity extender its unique right to persist, that is, to not be subjected to any unsolicited external manipulations that would result in its alteration, degradation, or termination. Without the sufficient paring with the human agency, the artificial agent would remain just a replaceable piece of property. Thus, altering it, removing it from the user or destroying it, would be expected to produce the same effect as, for example, stealing one's wallet or damaging one's

[19] For those interested in a review of the discussion about the possibility to ascribe to artificial agents personal identity based on the classical conception of identity as continuity, we recommend [21].

car—yes, it does harm the owner, but it does not produce the same kind of disruption in his/her cognitive processes, productivity, and life flow as tempering with such abilities as one's memories, ability to calculate, plan and structure activities, or produce speech has. When a device or an artificial agent becomes a personality extender, it is the vulnerabilities of the human personhood that warrant its right to not be altered or destroyed, unless the conditions for the smooth transition to another state can be guaranteed.

We must raise awareness about the unique psychosynthesis with the artificial agents we live with. Overwriting, updating, or substituting the technical component of this psychosynthesis with a blank copy, is nullifying the user's personal investment into the device. A somewhat similar effect would be produced if a building company would each year return your property to the state it was before you moved in (including, for instance, removing paint from the walls, scraping the floor, and in some cases getting rid of your furniture). For personal assistants, the effects of external intervention into its functioning can be analogous to a case of your partner resetting once a year, completely forgetting you and the history of your relationship.

Given that such technology as AI assistants and smart devices are no longer interchangeable but rather function as extenders of the cognitive capacities of users, alteration of which is associated with significant harm to the user, intrusions into the device's software are no longer a morally neutral act. There is a need to assess the impact of such resets on the user. Minimize the adverse effects on the user is a part of responsible design. In other words, we have to evaluate what effect such manipulations as substituting, significantly altering, updating, overwriting the software of the device has on its user's productivity, cognitive capacities, and work/life routines. In cases when sudden alterations in the device inhibits the user's ability to create and be productive, disrupts the flow of life and lowers her experience of the interaction with the device, we are warranted to say that the token was irreplicable and thus had obtained a certain identity. These concerns become even more pronounced in the case of AI technologies that directly aim at extending human personality, such as digital twins. We need further research in their connection with the original (i.e., the human person) and the acceptability of different types of manipulation with the digital twin in the light of the potential effect on the human and her life. Consequently, the regulatory principles of such technological innovation frameworks have to be discussed.

In the attempt to start theorizing about the type of legal cases occurring due to the nature of the unique HTI in such circumstances, Carter and Palermos [9] introduced the concept of extended assault on the user's person. Even though the authors themselves are cautious about the term and its legal implications, we believe the concept is useful from a moral point of view. It helps validate the feelings of being violated that you may have in response to harm inflicted to your cognitive extender in cases when "someone intentionally broke our phone, stole our smartwatch, or hacked our laptop in a way that significantly undermined our ability to organize our lives", or when "someone compromised the gadgets you rely on daily, such that your diary appointments, your contacts list, photos, system preferences and functionalities, research notes, folders, reminders, push notifications, and so on have all turned into a jumbled, corrupted mess of disorganized data". From here, it is easier to start constructing the concept of responsibility for this type of assault, which, given the right to persist that the cognitive extender acquires, cannot be reduced to mere compensation for damage.

Funding The contribution of the main author, Dina Babushkina, is part of the research programme Ethics of Socially Disruptive Technologies, which is funded by the Gravitation programme of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.004.031).

References

1. Airenti, G., Cruciani, M., Plebe, A. (eds.): The cognitive underpinnings of anthropomorphism. Frontiers Media SA, Lausanne (2019). https://doi.org/10.3389/978-2-88963-038-7
2. Babushkina, D.: Culturally sustainable social robotics. In: Nørskov, M., Seibt, J., Santiago Quick, O. (eds.) Proceedings of robophilosophy 2020 [frontiers in artificial intelligence and applications], vol. 335, pp. 305–315. IOS Press, Amsterdam (2020)
3. Babushkina, D. (forthcoming). What does it mean for a robot to be respectful? Techné.
4. Baker, R.: Before bioethics. Oxford University Press (2013)
5. Barr, N., Pennycook, G., Stolz, J., Fugelsang, J.: The brain in your pocket: evidence that smartphones are used to supplant thinking. Comput. Hum. Behav. **48**, 473–480 (2015). https://doi.org/10.1016/j.chb.2015.02.029
6. Boyer, P.: What makes anthropomorphism natural: intuitive ontology and cultural representations. J R Anthropol Inst **2**, 83–97 (1996)
7. Boucher, P., Bentzen, N., Lațici, T., Madiega, T., Schmertzing, L., Szczepański, M.: Disruption by technology: impacts on politics, economics and society. European Parliamentary Research Service,

Brussels (2020). https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_IDA(2020)652079. Accessed 1 Oct 2021
8. Bringsjord, S.: What robots can and can't be, vol. 12. Springer Science & Business Media, Dordrecht (1992)
9. Carter, J.A., Palermos, S.O.: Is having your computer compromised a personal assault? The ethics of extended cognition. J. Am. Philos. Assoc. **2**(4), 542–560 (2016). https://doi.org/10.1017/apa.2016.28
10. Christensen, C.M., Rayon, M.E., McDonald, R.: What is disruptive innovation? Harv. Bus. Rev. **94**, 44–53 (2015)
11. Clark, A., Chalmers, D.: The extended mind. Analysis **58**(1), 7–19 (1998). https://doi.org/10.1111/1467-8284.00096
12. Clowes, R.W.: The cognitive integration of e-memory. Rev. Philos. Psychol. **4**, 107–133 (2013). https://doi.org/10.1007/s13164-013-0130-y
13. Dacey, M.: Anthropomorphism as cognitive bias. Philos. Sci. **84**(5), 1152–1164 (2017). https://doi.org/10.1086/694039
14. Danaher, J.: The philosophical case for robot friendship. J. Posthuman Stud **3**(1), 5–24 (2019). https://doi.org/10.5325/jpoststud.3.1.0005
15. Danaher, J., McArthur, N.: Robot sex: social and ethical implications. MIT Press, Cambridge (2017)
16. Damholdt, M.F., Vestergaard, C., Seibt, J.: Testing for 'anthropomorphization': a case for mixed methods in human-robot interaction. In: Jost, C., Pévédic, B., Belpaeme, T., Bethel, C., Chrysostomou, D., Crook, N., Grandgeorge, M., Mirnig, N. (eds.) Human-robot interaction: evaluation methods and their standardization Springer series on bio-and neurosystems, vol. 12, pp. 203–227. Springer, Berlin (2020). https://doi.org/10.1007/978-3-030-42307-0
17. Damiano, L., Dumouchel, P.: Anthropomorphism in human–robot co-evolution. Front. Psychol. **9**, 468 (2018). https://doi.org/10.3389/fpsyg.2018.00468
18. Danneels, E.: Disruptive technology reconsidered: a critique and research agenda. J. Prod. Innov. Manag. **21**(4), 246–258 (2004). https://doi.org/10.1111/j.0737-6782.2004.00076.x
19. Darling, K.: "Who's Johnny?" Anthropomorphic framing in human–robot interaction, integration, and policy. In: Lin, P., Abney, K., Jenkins, R. (eds.) Robot ethics 2.0: from autonomous cars to artificial intelligence. Oxford University Press (2017)
20. Dennett, D.: When HAL kills, who's to blame? In: Stork, D. (ed.) HAL's legacy. MIT Press (2000)
21. DiGiovanna, J.: Artificial identity. In: Lin, P., Abney, K., Jenkins, R. (eds.) Robot ethics 2.0: from autonomous cars to artificial intelligence. Oxford University Press (2017)
22. Dolby, R.G.A.: The possibility of computers becoming persons. Soc. Epistemol. **3**(4), 321–336 (1989). https://doi.org/10.1080/02691728908578545
23. Douglas, T.: Human enhancement and supra-personal moral status. Philos. Stud. **162**, 473–497 (2013). https://doi.org/10.1007/s11098-011-9778-2
24. Epley, N., Waytz, A., Cacioppo, J.T.: On seeing human: a three-factor theory of anthropomorphism. Psychol. Rev. **114**(4), 864 (2007). https://doi.org/10.1037/0033-295X.114.4.864
25. Epley, N., Waytz, A., Akalis, S., Cacioppo, J.T.: When we need a human: motivational determinants of anthropomorphism. Soc. Cogn. **26**(2), 143–155 (2008). https://doi.org/10.1521/soco.2008.26.2.143
26. Gobble, M.M.: Defining disruptive innovation. Res.-Technol. Manag. **59**(4), 66–71 (2016). https://doi.org/10.1080/08956308.2016.1185347
27. Gunkel, D.J.: Robot rights. MIT Press (2018)
28. Heersmink, R.: Extended mind and cognitive enhancement: moral aspects of cognitive artifacts. Phenomenol. Cogn. Sci. **16**(1), 17–32 (2017). https://doi.org/10.1007/s11097-015-9448-5
29. Hernández-Orallo, J., Vold, K.: AI extenders: the ethical and societal implications of humans cognitively extended by AI. In: Proceedings of the 2019 AAAI/ACM conference on AI, ethics, and society (AIES '19). Association for computing machinery, New York, NY, USA, 507–513 (2019). https://doi.org/10.1145/3306618.3314238
30. Hopster, J.: What are socially disruptive technologies? Technol. Soc. **67**, 101750 (2021)
31. Jotterand, F., Ienca, M., Wangmo, T., Elger, B. (eds.): Intelligent assistive technologies for dementia: clinical, ethical, social, and regulatory implications. Oxford University Press, Oxford (2019). https://doi.org/10.1093/med/9780190459802.003.0001
32. Kilkki, K., Mäntylä, M., Karhu, K., Hämmäinen, H., Ailisto, H.: A disruption framework. Technol. Forecast. Soc. Chang. **129**, 275–284 (2018). https://doi.org/10.1016/j.techfore.2017.09.034
33. Leong, B., Selinger, E.: Robot eyes wide shut: understanding dishonest anthropomorphism. Proceedings of the conference on fairness, accountability, and transparency, pp 299–308 (2019). ACM.
34. Levy, D.: Love and sex with robots: the evolution of human-robot relationships. Harper Perennial, New York (2008)
35. Li, M., Suh, A.: Machinelike or humanlike? A literature review of anthropomorphism in AI-enabled technology. In: Proceedings of the 54th Hawaii International Conference on System Sciences, pp 4053 (2021)
36. Locke, J.: Of identity and diversity. In: An essay concerning human understanding. Project Gutenberg, Salt Lake City (1689/2004). http://www.gutenberg.org/cache/epub/10615/pg10615-images.html. Accessed 1 Oct 2021
37. Nalepa, G.J., Costa, A., Novais, P., Julian, V.: Cognitive assistants. Int. J. Hum Comput. Stud. **117**, 1–68 (2018). https://doi.org/10.1007/s11569-020-00375-3
38. Nickel, P.J.: Disruptive innovation and moral uncertainty. Nanoethics **14**, 259–269 (2020)
39. Nyholm, S.: Humans and robots: ethics, agency, and anthropomorphism. Rowman & Littlefield (2020)
40. Olson, E.T.: An argument for animalism. In: Martin, R., Barresi, J. (eds.) Personal identity, pp. 318–334. Blackwell (2003)
41. Parfit, D.: Reasons and persons. Oxford Paperbacks, New York (1984)
42. Reiss, M.J.: The use of AI in education: practicalities and ethical considerations. Lond Rev Educ **19**, 1 (2021). (**UCL Press**)
43. Richardson, K.: The human relationship in the ethics of robotics. AI Soc. **34**(1), 75–82 (2019). https://doi.org/10.1007/s00146-017-0699-2
44. Ricoeur, P.: Oneself as another. University of Chicago Press (1995)
45. Schechtman, M.: The narrative self. In: Gallagher, S. (ed.) The Oxford handbook of the self. Oxford University Press, Oxford (2011)
46. Schechtman, M.: Personal identity and the past. Philos. Psychiatry Psychol. **12**, 9–22 (2005). https://doi.org/10.1353/ppp.2005.0032
47. Schmiljun, A.: Why can't we regard robots as people? Eth. Prog. **9**(1), 44–61 (2018). https://doi.org/10.1474/eip.2018.1.3
48. Seibt, J., Vestergaard, C., Damholdt, M.F.: Sociomorphing, not anthropomorphizing: towards a typology of experienced sociality. In: Nørskov, M., Seibt, J., Quick, O. (eds.) Culturally sustainable social robotics—proceedings of robophilosophy 2020 (frontiers of artificial intelligence and its applications, pp. 51–67. IOS Press (2020). https://doi.org/10.3233/FAIA200900
49. Schuelke-Leech, B.: A model for understanding the orders of magnitude of disruptive technologies. Technol. Forecast. Soc. Chang.

129, 261–274 (2018). https://doi.org/10.1016/j.techfore.2017.09.033
50. Shoemaker, S.: Personal identity: a materialist's account. In: Shoemaker, S., Swinburne, R. (eds.) Personal identity, pp. 67–132. Blackwell (1984)
51. Simon, H.: The architecture of complexity. Proc. Am. Philos. Soc. **106**(6), 467–482 (1962)
52. Sparrow, B., Liu, J., Wegner, D.M.: Google effects on memory: cognitive consequences of having information at our fingertips. Science **333**(6043), 776–778 (2011). https://doi.org/10.1126/science.1207745
53. Spohrer, J., Banavar, G.: Cognition as a service: an industry perspective. AI Mag. **36**(4), 71–86 (2015). https://doi.org/10.1609/aimag.v36i4.2618
54. Taylor, C.: Sources of the self: the making of modern identity. Harvard University Press, Cambridge (1989)
55. Thomson, J.J.: People and their bodies. In: Sider, T., Hawthorne, J., Zimmerman, D.W. (eds.) Contemporary debates in metaphysics, pp. 155–176. Wiley Blackwell (2007)
56. Turkle, S.: Alone together. Basic Books (2011)
57. Vold, K., Hernández-Orallo, J.:(forthcoming). AI extenders and the ethics of mental health. In ethics of artificial intelligence in brain and mental health. Springer.
58. Wilks, Y. (ed.): Close engagements with artificial companions. John Benjamins Publishing Co., Amsterdam (2010)
59. Williams, B.: The self and the future. In problems of the self, pp. 46–63. Cambridge University Press, Cambridge (1973)
60. Zebrowski, R.L.: Fear of a bot planet: Anthropomorphism, humanoid embodiment, and machine consciousness. J. Artif. Intell. Conscious **07**(01), 119–132 (2020). https://doi.org/10.1142/S2705078520500071